Title: Dependence analysis of categorical data from banking

Author: Miroslav Khýr

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Jitka Zichová, Dr., Department of Probability and Mathematical Statistics

Abstract: The aim of this work is describing in detail the theory of the log - linear expansion and graphical models for random vectors with a discrete distribution. Such vector can be used for modeling categorical variables for example in a population of borrowers by a bank . We show how to estimate the probability of an individual category. We use a log - likelihood function. Independence graph can represent conditional independence of discretely distributed random variables. Using this theory, especially using deviance as test statistics, we can examine whether same data correspond to the selected graphical model. At the end of this work we apply the described theory to real data and determine the graphical model best fitting the dependence structure in a database from banking. From this graph we can deduce which variables are dependent and which are independent.

Keywords: Log - linear expansion, graphical model, log - likelihood function ,deviance.